Long-term 3D Registration Method Based on LCT Tracking and Improved ORB Detection

Jiu Yong[1, 2*], Yangping Wang[1, 2†], Xiaomei Lei[3†], Fang Yong[4†], Zhenhai Zhang[1, 2, 5†]
1. Lanzhou Jiaotong Univeristy, Lanzhou, China
2. Gansu Provincial Engineering Research Center for Artificial Intelligence and Graphics & Image Processing, Lanzhou China
*E-mail: **1527983509@qq.com** Tel: +86-18794844571
†E-mail: 1328396793@qq.com Tel: +86-13519311970
3. Gansu Meteorological Service, Lanzhou 730020,China
†E-mail: 1530194583@qq.com Tel: +86-18293106629
4. Sichuan University, Chengdu, China
†E-mail: 2936542988@qq.com Tel: +86-13281261083
5. Lanzhou Bocai Technology Co., Ltd., Lanzhou, China
†E-mail: 764411629@qq.com Tel: +86-18215188963

Abstract—**Aiming at the complex environment such as fast moving of registered area, occlusion, illumination change and the high requirement of real-time and precision of feature detection in the 3D registration of augmented reality system, a long-term 3D registration method based on LCT tracking and improved ORB detection is proposed in this paper. Firstly, the reliability of LCT algorithm in long-term tracking is used to track the area to be registered in augmented reality; secondly, ORB algorithm with excellent real-time performance is improved by setting adaptive thresholds, number of feature points and distance thresholds to optimize the dense area of image feature points. Parallel algorithm is used to retain feature points with larger eigenvalues, and discrete difference feature is used to enhance illumination unevenness. The stability of ORB operator under uniform change can solve the problem of low precision of feature detection and poor anti-jamming ability. Finally, the 3D registration matrix is calculated by using the detected feature points to enhance the real world. The simulation results show that the LCT algorithm has high reliability in long-term tracking and registration. Compared with ORB algorithm, the improved ORB algorithm improves the precision of feature detection by about 22%. It effectively improves the real-time and precision of feature matching in augmented reality system. The performance of the long-term 3D registration method based on LCT tracking and improved ORB detection is excellent, which improves the robustness, stability and practicability of the augmented reality system.**

Keywords: LCT Tracking; ORB Feature Detection; Augmented Reality; 3D Registration; Adaptive Threshold

I. INTRODUCTION

Augmented Reality (AR) is the superposition of virtual information generated by computer and the real world that people see, giving people the effect of visual enhancement. AR technology has the characteristics of combination of virtual reality, real-time interaction and 3D registration. 3D registration technology is the key, foundation and difficulty in the construction of AR system. So the quality of 3D registration technology is directly related to the performance of AR system. At present, the mainstream 3D registration method of AR system is tracking registration method based on visual image[1-2].

At present, in the field of computer vision target tracking, how to ensure the robustness and speed of the algorithm is a great challenge for target tracking, when the target area is tracked for a long time, it is easy to change, move quickly or occlude. In recent years, the research of target tracking algorithms represented by correlation filtering has developed rapidly[3]. The correlation filtering can determine the central position of the tracking target by searching the maximum response, and has excellent speed performance in target tracking. In 2010, Bolme et al. first applied correlation filtering to target tracking and proposed the MOSSE (Minimum Output Sum of SquaredError) algorithm[4]. In 2012, Heriques et al. introduced correlation filtering into the core space and used cyclic shift to conduct intensive sampling. This method became the standard of correlation filtering. The next development is mainly embodied in improving CSK (Exploiting the Circulant Structure of Tracking-by-detection with Kernels) algorithm in feature selection and application[5]. In 2015, Henriques et al. subsequently proposed KCF (Kernelized Correlation Filters) algorithm by combining CSK algorithm with HOG (Histogram of Oriented Gradient) feature[6]. In 2017, Danelljan et al. put forward ECO algorithm by incorporating depth feature into correlation filtering, and achieved good tracking effect[7]. However, the correlation filtering method is sensitive to fast deformation and motion, and the tracking effect is not good. In addition, 3D registration of AR systems usually requires long-term tracking for the target, and long-term tracking needs to ensure that the model is long-term effective, that is, compared with short-term tracking, it is more vulnerable to rapid target movement, changes in the ratio of width to height outside the field of vision and complete occlusion[8]. In 2011, in order to solve this problem, Kalal et al. combined tracking with detection by using re-detector to prevent tracking failure and the results of the detector to train the tracker, and proposed a TLD (Tracking-Learning-Detection) algorithm[9-10]. However, the tracking speed of this method is very slow, and the effect is not good when the target changes greatly. In 2015, Ma Chao and others synthetically considered the context and scale transformation of the target. Based on the correlation filtering and the re-detector based on random fern, they proposed LCT (Long-term Correlation Tracking) algorithm[11]. By adding detection mechanism, the LCT algorithm performs well in occlusion and out-of-sight of the target and has high tracking precision.

At present, SIFT (Scale Invariant Feature Transform) algorithm[12], SURF (Speeded Up Robust Features) algorithm[13] and ORB (Oriented FAST and Rotated BRIEF) algorithm[14] are popular image feature detection methods. The SIFT algorithm proposed by David Lowe et al. in 1999 has strong matching ability. It can keep unchanged for the image scale transformation, rotation transformation and brightness change. It also has strong stability for noise pollution, perspective transformation and affine transformation. However, the speed of feature detection is slow. The SURF algorithm proposed by Herbert Bay and others in 2008 according to the SIFT algorithm has high recognition rate of feature points, good robustness in light, perspective and scale changes, and to a certain extent solves the shortcomings of the SIFT algorithm in computing time. The ORB algorithm proposed by Ehtan Ruble et al. in 2011 is essentially a combination of FAST corner detection algorithm and BRIEF binary feature description algorithm, and ORB algorithm is one order of magnitude faster than SURF and two orders faster than SIFT in matching time, but it has the problems of low precision of feature detection and poor anti-interference ability makes the registration precision of AR system vulnerable to image texture and uneven illumination.

Based on the above research, this paper proposes a long-term 3D registration method based on LCT tracking and improved ORB detection. The registration area is tracked by LCT algorithm with excellent long-term tracking performance. Secondly, the registration area is detected by improved ORB algorithm. The algorithm enriches the number and distance of feature points in the region by setting adaptive threshold. Finally, feature matching is carried out to calculate the 3D registration matrix to realize long-term tracking and registration of the real world.

II. LCT Long-Term Target Tracking Algorithms

LCT tracking algorithm aims at long-term target tracking, and add scale estimation on the basis of kernel correlation filtering of HOG features. Random fern algorithm is used for re-detection to ensure that the target can be precisely tracked after occlusion, so that the tracking precision of the algorithm can be maintained for a long time without tracking drift.

A. Kernel Correlation Filtering

LCT tracker is divided into 3 parts: position filter R_c, confidence filter R_t and scale filter R_s, which are used to track target location, confidence confirmation and target scale estimation respectively. They are all related filtering methods.

Choose the image block x which contains the target and part of the background area. The size of the block is $M \times N$ to create samples $x_{m,n}$, $(m,n) \in \{0,1,,\cdots,M-1\} \times \{0,1,\cdots,N-1\}$ by cyclic shifting in the range. Establish the ridge regression model of the objective function as follows:

$$\min \sum_{m,n} \| \phi(x_{m,n}) \cdot w - y(m,n) \|^2 + \lambda \| w \|^2 \tag{1}$$

In the formula, λ is the parameter of regularization term; $y(m,n)$ is the tag set by Gaussian function, and the maximum value is at the target center point. $\phi(x_{m,n})$ is the mapping in the kernel space, and the Gaussian kernel is chosen for the kernel function.

By using fast Fourier transform to calculate correlation, the target solution can be expressed as $w = \sum_{m,n} a(m,n)\phi(x_{m,n})$, in which correlation coefficient a is defined as follows:

$$A = F(a) = \frac{F(y)}{F(\phi(x) \cdot \phi(x))} \tag{2}$$

In the formula, F represents discrete Fourier transform. In the new frame, the response output in search window z is:

$$\hat{y} = F^{-1}(A \odot F(\phi(z) \cdot \phi(\hat{x}))) \tag{3}$$

In the formula, $\hat{x}$ represents the learned object appearance model and $\odot$ represents the element dot product. The maximum value of $\hat{y}$ response is new location of the target.

B. Model Updating and Re-detection

In order to ensure the adaptability of the model to the new appearance of the target, it is necessary to update the model. The updated formula is as follows:

$$\hat{x}^t = (1-\alpha)\hat{x}^{t-1} + \alpha x^t \tag{4}$$

$$\hat{A}^t = (1-\alpha)\hat{A}^{t-1} + \alpha \tag{5}$$

In the formula, t represents the current frame and α is the learning rate.

After occlusion or out of view occurs in the tracking target area, it is easy to cause tracking failure, and too much background information will lead to the reduction of the precision of the model. In order to reduce the adverse effects caused by this situation, LCT algorithm adopts two control methods of confidence analysis and re-detection in the tracker[10].

III. Improved ORB Feature Detection Algorithm

ORB algorithm is essentially the combination of FAST corner detection algorithm and BRIEF binary feature description algorithm, and ORB algorithm has excellent real-time performance, but it has the problems of low precision of feature detection and poor anti-interference ability. This paper sets adaptive threshold by ORB algorithm, and makes use of the number and distance relationship of feature points in rich regions of image texture and discrete difference features. To improve the precision of feature detection and anti-interference ability of ORB operator, solve the problem that the registration precision of virtual and real images is easily affected by image texture and uneven illumination, and ensure the precision and real-time requirements of registration precision.

A. ORB Feature Detection Algorithm

ORB algorithm is a feature point detection and description algorithm based on visual information proposed by Ruble et al. on ICCV. It uses FAST to detect feature points, adds direction information of FAST features, and uses BRIEF to describe feature points. It also improves the disadvantages of BRIEF that it does not have rotation invariance and is sensitive to image noise[14]. ORB feature extraction can be divided into two parts: (1) FAST interest point detection with direction; (2) BRIEF interest point description with rotation invariance.

(1) Interest Point Detection

The FAST feature points are detected by using segmentation detection criterion. If the radius of the circle is r and the center of the pixel is p, there are r connected pixels I_k (where $k = 1, 2, \cdots, n$). The following formula can be used to determine whether the pixel p is a corner:

$$CRF = \begin{cases} 1 & if \; |I_p - I_k| > t \\ 0 & else \end{cases} \tag{6}$$

In the formula, I_k is the gray value of any pixel; the gray value of p is I_p, and t is a small known threshold. If the number of CRF=1 is more than a given threshold t', the point is considered as a candidate point, usually $t' = 12$. As shown in Fig. 1.

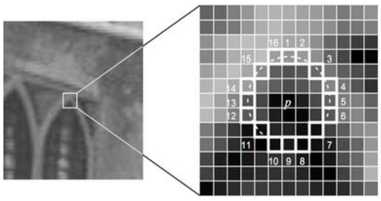
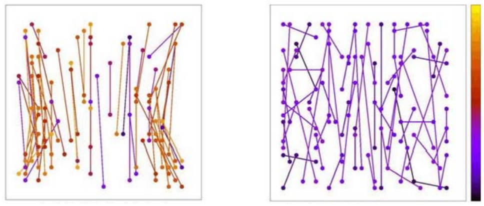


Fig. 1 Schematic diagram of ORB feature detection.

By building image pyramids and introducing scale characteristics and removing the direction of top feature points by gray centroid method[15], make FAST interest points have direction, and BRIEF operator is extracted according to this direction.

(2) Generating Operators with Rotation Invariance

In order to solve the problem of noise sensitivity, ORB algorithm uses Gauss kernel filter to preprocess the image, and two methods are used to solve the problem that BRIEF operator does not have rotation invariance. 1) Control the direction of FAST feature points; (2) Use the greedy exhaustive algorithm to find pairs of random points with low correlation, that is, the correlation coefficient is close. 0.5 image point pair; In Fig.2, Figure (a) is a BRIEF feature. After introducing the main direction, the correlation of random point pairs becomes larger because of the change of the main direction of feature points. Figure (b) is the result of reducing the correlation of feature points after using greedy exhaustive algorithm. The right color bar is the correlation of each test feature point, and the lighter the color, the higher the correlation.



(a) (b)

Fig. 2 ORB Operators with Rotation Invariance

B. Improvement of ORB algorithm

To solve the problem of poor precision and anti-interference ability of traditional ORB feature detection method, the ORB algorithm is improved as shown in Figure 4. In the process of feature detection, an adaptive corner detection method is added to optimize the dense area of image feature points by setting the number and distance threshold of ORB feature points. Parallel algorithm is used to retain feature points with larger eigenvalues, and then discrete difference feature is used to enhance the stability when illumination changes unevenly, so as to improve the precision and anti-interference ability of ORB feature detection.

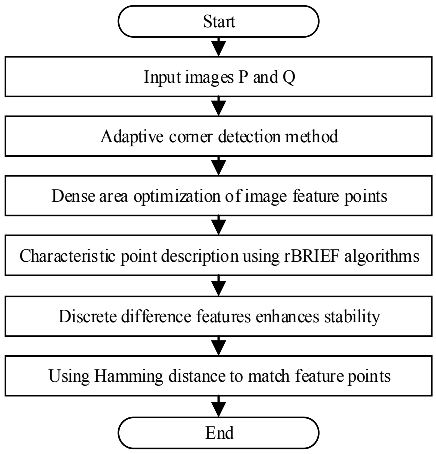


Fig. 3 ORB algorithm improvement process

(1) Adaptive Corner Detection

In the process of corner detection, FAST algorithm calculates the difference between the gray value of the candidate corner and 16 points around it. If the number of pixels satisfying formula (1) exceeds 9 or 12, the candidate is considered as a corner. The selection of threshold θ determines the number of feature points. The larger the threshold θ is, the fewer features can be extracted in the image with low contrast. Therefore, different θ values should be selected according to different contrast images, while the θ values of traditional algorithms depend on artificial settings, which cannot be applied to different contrast images, and the anti-jamming performance is poor.

In order to solve the above problems, an adaptive threshold is set to replace the artificial setting in the original algorithm. Through the analysis of image contrast and gray value, an adaptive selection method of threshold θ under different image contrast is proposed, so that the value of threshold θ can change according to the difference of image contrast.

$$\theta = \alpha \bullet (\frac{1}{n}\sum_{i=1}^{n}(I(x_i) - I(\overline{x}))^2) \tag{7}$$

In the formula, α is the scale factor, which determines the number of corners detected; $I(x_i)$ is the gray value of each pixel in the image; $I(\overline{x})$ is the gray mean of the image. In the follow-up experiments of this paper, through comparative analysis, the use of $\alpha = 0.01$ is more reasonable, which guarantees the number of corners and restrains the generation of pseudo-corners.

(2) Dense Area Optimization of Image Feature Points

ORB algorithm calculates the Harris interest value of FAST local region in order to eliminate a large number of feature points extracted from the edge of FAST corner, and ranks them according to the size of interest value[14]. The sorting time of original interest ranking algorithm is stable in a certain range of interest points, but when the number of features exceeds a certain threshold, the sorting time increases abnormally, which significantly affects the ORB feature detection time. Moreover, the number of extracted features increases significantly in regions rich in texture. On the one hand, the time of feature description and matching will increase significantly, which reduces the efficiency of feature point matching; on the other hand, the number of feature points is too large, and the feature points in dense areas of the image are difficult to identify, which is prone to mismatch, resulting in the decline of the precision of feature point matching[16].

To solve the above problems, the strategy of image segmentation and neighboring feature points removal is adopted to ensure the uniformity of feature points distribution and avoid feature points clustering. Before feature point screening, the image is divided into blocks[17] and the total number of feature points is set to N_{thd} to ensure the uniformity of the distribution of feature points and the stability of the total number of feature points. The feature points detected in each image block are sorted according to the size of the feature values, and N_{thd} points with larger feature values are selected. In this way, the significant feature points in each region of the image have been detected and retained, and the uniformity of the distribution of the detected feature points has been achieved. However, clustering of feature points may occur in local regions, which may lead to a significant increase in the probability of mismatching when matching feature points. This phenomenon can be eliminated by setting the nearest neighbor distance D_{thd}.

The number of feature points $N_{thd} = 300$ and normalized distance $D_{thd} = 0.02$ [17] are set for different images. $N_{thd} = 300$ is obtained on the premise of weighing registration precision and registration time, and the threshold can be modified for different situations; $D_{thd} = 0.02$ ensures that the registration can be successful even if the feature points are close to each other in the texture-rich region, that is, the threshold of near-neighbor distance is set. When the total number of local Harris interest points extracted by FAST operator is greater than N_{thd}, the feature points are balanced by removing the densely distributed points, and then a parallel sorting algorithm based on OpenMP parallel programming model is designed to quickly sort Harris and retain the appropriate feature points. The specific flow of the algorithm is as follows:

Step1. Set the number of uniform blocks n as CPU core number, number threshold N_{thd} and distance threshold D_{thd}, and mark $FirstInvoke = 1$ as the feature point with the largest eigenvalue.

Step2. Obtain the Harris point detected in the ORB algorithm flow and store it in container $vector < KeyPt > Vecpts$.

Step3. $GetSizeOfPts(VecPts)$ obtains the total number of feature points P_{sum}, if $P_{sum} < N_{thd}$ outputs $VecPts$, jump to Step6;

Step4. $GetMaxOfPts(VecPts)$ obtains the maximum feature point $VecPts$ of the current remaining feature points of $P_{\max}$, if $FirstInvoke == 1$, then $P_{pre\_\max} = P_{\max}$;

Step5. If $GetDist(P_{\max}, P_{pre\_\max}) > D_{thd}$, delete $P_{\max}$ from $Vecpts$, $\mathrm{Re}move(P_{\max}, VecPts)$, $P_{pre\_\max} = P_{\max}$, then jump to Step3 to continue execution, otherwise jump to Step4 to continue execution;

Step6. Implement a parallel and fast sorting algorithm of Harris interest value based on OpenMP, and retain the appropriate feature points. Input the sorted Harris feature dataset $VecPts$ and size $count$, and output the sorted Harris feature dataset $VecPts$. The specific process is as follows:

(a) Parallel execution code segments are generated by OpenMP compiler instructions;

(b) Each thread obtains its own thread number $threadNo$ and total thread number $threadTotal$;

(c) Each thread calls the sorting algorithm and executes the parallel quick sorting function $\mathrm{P}arallelQuicksort()$;

(d) Call $merge()$ function to merge part of the ordered data generated by threads.

(3) Discrete Difference Features Enhance Stability

In ORB feature extraction, in order to solve the problem of noise sensitivity of BRIEF feature point algorithm, the binary string is obtained by comparing several pixel blocks with integral graph, which solves the problem of noise to a certain extent. However, experiments show that ORB algorithm is still sensitive to the uneven change of illumination. Nowadays, noise is mainly reduced by filtering operation, but the time consumed by filtering operation reduces the real-time performance of the system. In this paper, by using discrete difference features to enhance its stability, 31×31 sub-windows in 5×5 pixel neighborhoods are compared with several pairs of pixel blocks using integral graphs, and a binary string is obtained. Test τ is defined as follows:

$$\tau(p;x,y)=\begin{cases}1, & |p(x)-p(y)|>\theta' \\ 0, & other\end{cases} \tag{8}$$

In the above formula, the gray difference $|p(x)-p(y)|$ between image blocks is used as the criterion of τ. When the illumination and background change unevenly, the gray difference $|p(x)-p(y)|$ has certain stability. Therefore, the gray difference $|p(x)-p(y)|$ is used as the criterion to construct the improved ORB operator. At the same time, the threshold θ' can take the value of $70\sim 80$ [18] in different images, and the average value $\theta'=75$ in this paper.

IV. Computation of 3D Registration Matrix

After completing the feature detection, the method of feature matching in this paper adopts the Hamming distance matching method. Its principle is to calculate the number of different characters in the corresponding positions in two equal-length strings. After matching video frames, the best matching point pairs are selected by using ML-RANSAC (Multilevel Random Sample Consensus) algorithm[19]. The ML-RANSAC algorithm is developed to robustly estimate velocity and position of the multiple moving objects in an unknown environment whereas the state of the objects (static or dynamic) is not known a priori. The main characteristic of the algorithm is its ability to address both static and dynamic objects and to detect and track moving objects without dividing the problem into two separate parts. Using the results of target tracking and improved ORB algorithm feature detection and matching, the 3D registration matrix is calculated to enhance the real world. In this paper, the initial 3D registration matrix is obtained by using the first frame image, and then the subsequent 3D registration matrix is obtained by using the improved ORB algorithm's feature points and the initial registration matrix.

A. 3D Registration Projection Transformation Relations

Considering the coordinates of feature points, camera and display screen projection, the relationship is shown in Fig.4.

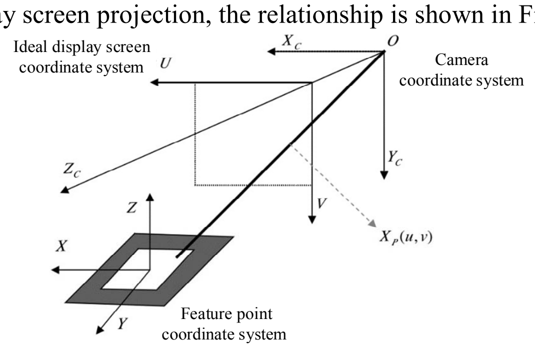


Fig. 4 Definition of coordinate system

In Fig.4, $O-XYZ$ is the feature point coordinate system, $O-X_cY_cZ_c$ is the camera coordinate system, UV is the display screen coordinate system, and $X_P(u,v)$ is the projection of any point on the special graphics on the display screen coordinate system UV.

The coordinate transformation relationship between any point $P(X,Y,Z)$ on the feature graph and the projection corresponding point $X_P(u,v)$ on the display screen is satisfied[20]:

$$h\begin{bmatrix}u\\v\\1\end{bmatrix}=\begin{bmatrix}s_xf & 0 & u_0 & 0\\ 0 & s_yf & v_0 & 0\\ 0 & 0 & 1 & 0\end{bmatrix}\begin{bmatrix}X_C\\Y_C\\Z_C\\1\end{bmatrix} \tag{9}$$

In the formula, f is the focal length of the camera, s_x and s_y are the scale factors, u_0 and v_0 are the coordinates of the center of gravity of the image. The coordinate system defined above and the projection transformation relation can be obtained:

$$\begin{aligned}&[x_c,y_c,1]^T\\&=\lambda CT_m[X_m,Y_m,Z_m,1]^T\\&=\lambda C[R_1,R_2,R_3,T][X_m,Y_m,Z_m,1]^T\end{aligned} \tag{10}$$

In the formula, $[x_c,y_c]$ is the coordinate of point X in ideal screen coordinate system, $[X_m,Y_m,Z_m]$ is the coordinate of point X in characteristic graphics coordinate system, C is the internal parameter matrix of unknown camera, T_m is the 3D registration matrix needed for 3D registration.

B. Feature Matching

The initial 3D registration matrix $T_m^{'}$ and camera parameter matrix C are acquired through the target tracking parameters and projection transformation relations of the first frame, and then the scene real-time image is acquired. The improved ORB algorithm is used to detect feature point (x_c,y_c) in the area be registered. Then the positive correlation operation is carried out between each extracted feature point and the initial frame feature information stored in the system. Selecting the most relevant feature points from the initial feature points set as the matching feature points corresponding to the feature points of the current region to be registered. The coordinates of the corresponding points in the feature graphics coordinate system are $[X_m,Y_m,Z_m]$. The calculation method of positive correlation degree is as follows:

$$S=\frac{\sum_{i=1}^{N}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{N}(x_i-\overline{x})^2}\sqrt{\sum_{i=1}^{N}(y_i-\overline{y})^2}} \tag{11}$$

In the formula, x_i is the image block pixel value, $\overline{x}$ is the average value of the image block pixel, y_i is the template pixel value, and $\overline{y}$ is the average value of the template pixel.

C. Computing 3D Registration Matrix

After feature template matching step, a set of candidate feature points (X_{ci}, X_{mi}) ($1 < i <$ matching feature points logarithm) can be obtained, and the 3D registration matrix can be calculated by using RANSAC algorithm. The specific calculation steps are as follows:

Step1. Four pairs of non-collinear feature points are randomly selected from the set of candidate feature points.

Step2. Use formula (10) to calculate 3D transformation matrix as candidate 3D registration matrix.

Step3. According to the candidate 3D registration matrix and feature set $T_m(X_{ci}, X_{mi})$, the distance d_i between X_{ci} and $\lambda C T_m X_{mi}$ corresponding to each pair of matching feature points is calculated. The threshold is set to 2 pixels, and T_m whose distance is less than or equal to the threshold is selected as the 3D registration matrix T_m corresponding to the current image.

Step4. Calculate the matrix T_m, and then use the matrix T_m to register the current frame image in 3D.

V. EXPERIMENTAL RESULTS AND ANALYSIS

In order to verify the long-term 3D registration method of AR based on LCT tracking and improved ORB detection, seven sets of video sequence data (Car24, Doll, Girl2, Human3, Liquor, RedTeam, Rubik) with more than 1500 frames in the visual tracking reference datasets[21] are selected and tested on the hardware and software platform shown in Tab.1. Compare LCT algorithm with TLD and CSK in precision, success rate and efficiency of target tracking, and then compare improved ORB algorithm with ORB algorithm and SURF algorithm in error and rate of feature detection, finally calculate 3D registration matrix by feature matching, superimposing OpenGL cubes to enhance the real world.

Tab.1 Experimental software and hardware platform

Hardware platform	Software platform
	(1) Windows10 OS
(1) PC (memory: 16G, CPU:Intel i7-8700)	(2) Programing language：C++
(2) Lenovo display device	(3) VS2015+OpenCV3.4.0
(3) Camera	(4) ARToolKit+OpenGL
	(5) MatLab2015a

A. Results and Analysis of LCT Target Tracking Algorithms

In the complex environment of the long-term sequence and Illumination Variation, Scale Variation, Occlusion, Motion Blur, Fast Motion, Out-of-Plane Rotation, Out-of-View and Background Clutters, target tracking experiments were carried out on 7 groups of video sequences, and the tracking results were compared and analyzed.

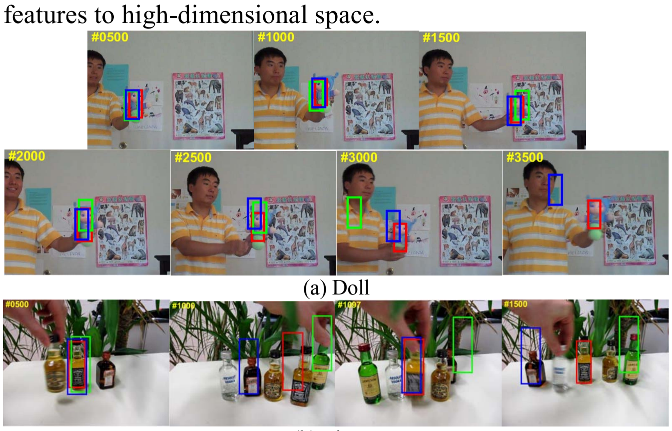
(1) Tracking Results in Complex Environments

Seven sets of data including Car24、Doll、Girl2、Human3、Liquor、RedTeam、Rubik grades were selected for experiment. Tab. 2 shows the main challenges of video sequences and the number of frames.

Tab.2 Test video in the experiment

Video	Frames	Challenges
Car24	3059	Illumination Variation, Scale Variation, Background Clutters
Doll	3872	Scale Variation, In-Plane Rotation, Out-of-Plane Rotation
Girl2	1500	Scale Variation, Occlusion, Deformation, Motion Blur, Out-of-Plane Rotation
Human3	1698	Scale Variation, Occlusion, Deformation, Out-of-Plane Rotation, Background Clutters
Liquor	1741	Illumination Variation, Scale Variation, Occlusion, Motion Blur, Fast Motion, Out-of-Plane Rotation, Out-of-View, Background Clutters
RedTeam	1918	Scale Variation, Occlusion, In-Plane Rotation, Out-of-Plane Rotation, Low Resolution
Rubik	1997	Scale Variation, Occlusion, In-Plane Rotation, Out-of-Plane Rotation

Seven sets of data in Tab.2 are tested on MatLab+OpenCV platform. LCT algorithm, CSK algorithm and TLD algorithm are evaluated with default parameters of source code. The experimental results of longest Doll dataset (3872 frames) and the most complex Liquor dataset (Illumination Variation, Scale Variation, Occlusion, Motion Blur, Fast Motion, Out-of-Plane Rotation, Out-of-View, Background Clutters) are shown in Fig.5. The experimental results of Fig. 5 show that LCT algorithm performs well in long-term tracking and complex environment, but when the Liquor data set is 1000 frames, the tracking error of LCT algorithm appears, which gradually decreases in subsequent tracking, and can precisely track the target position after 97 frames.TLD algorithm and CSK algorithm cannot track the target position at 1000 frames of Liquor dataset. TLD algorithm can track the target position precisely after 97 frames, but CSK algorithm cannot track the target position. That is, the tracking effect of CSK algorithm is poor in the case of fast jitter, mainly because it failed to obtain a more discriminant regression function when mapping features to high-dimensional space.



(a) Doll

(b) Liquor

LCT TLD CSK

Fig.5 Experimental results of target tracking

(2) Evaluation and Analysis of Algorithms

Quantitative analysis of the algorithm is to initialize the algorithm according to the exact position in the first frame, then run the algorithm in a test sequence, and finally get the results of average precision, success rate and efficiency, namely OPE (One-Pass Evaluation). The definitions of precision, success rate and efficiency are as follows:

(a) Precision rate. The standard widely used in tracking precision evaluation is the center position error, which is the average Euclidean distance between the center position of tracking target and the precise position of manual calibration. However, when tracking lost target, the tracking position of the output is random, and the average error value may not be able to precisely estimate the tracking performance. In this paper, the precision is the ratio of the number of frames to the total number of frames within the threshold φ of the given precision value at predicted position, where the error threshold $\varphi = 20px$ [22]. The evaluation results of precision are shown in Fig. 6.

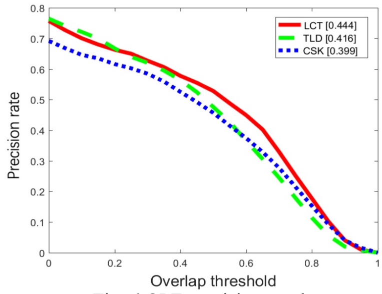


Fig. 6 OPE precision results

(b) Success rate. The evaluation criterion of success rate is the overlap rate of borders. Assuming the tracking boundary box τ_t and the precision boundary box τ_a, the repetition rate is defined as $S = \frac{|\tau_t \cap \tau_a|}{|\tau_t \cup \tau_a|}$, where $\cap$ and $\cup$ represent the intersection and union of two regions respectively, and $|\cdot|$ represents the number of pixels in the region. In order to evaluate the performance of the tracking algorithm in the video sequence, the number of successful frames of $S > t_0$ is calculated. The success rate gives the ratio of the number of successful frames when the threshold varies between 0 and 1, and gives a threshold (e.g.). t_0 = 0.5), but it is unfair or unrepresentative to use success rate for tracking evaluation. Therefore, this paper uses AUC (area under curve) of each success rate to evaluate the tracking algorithm. The evaluation results of success rate are shown in Fig. 7.

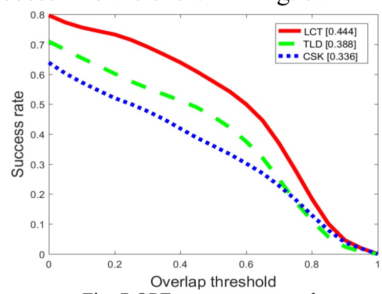


Fig. 7 OPE success rate results

(c) Efficiency. For the 7 groups of data used, the average FPS of each algorithm in OPE is shown in Fig. 8.

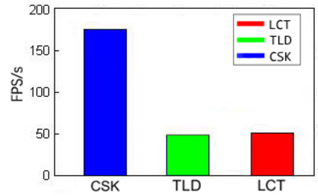


Fig.8 Comparison of average running speed of 3 algorithms

As shown in Fig. 6, Fig. 7 and Fig. 8. in the quantitative analysis of 7 groups of video sequences, the precision evaluation results and success rates of some tracking algorithms are different because different indicators are used to measure the different characteristics of tracking algorithms, and the AUC value is more precision than a threshold value in the overall evaluation of tracking algorithms. Therefore, in the follow-up discussion, the success rate is the main evaluation basis, and precision is only used as a supplement. Compared with TLD algorithm and CSK algorithm, LCT algorithm performs better in long-term tracking and complex environment because LCT algorithm adds a third correlation filter responsible for detecting target confidence on the basis of a translation correlation filter and a scale correlation filter, which makes LCT algorithm have higher precision and success rate.

B. Detection Result and Analysis of Target Position Feature Based on Improved ORB Algorithm

The target position tracked by LCT algorithm is detected by ORB algorithm and improved ORB algorithm respectively. At the threshold of $t' = 12$, the feature detection results of Doll data set and Liquor data set at 1500 frames are shown in Fig.9. Compared with the experimental results of the ratio factor of α, which are 0.01, 0.02 and 0.03, $\alpha = 0.01$ is more reasonable, guaranteeing the number of corner points, while restraining the generation of pseudo-corner points, at the same time, the number of feature points $N_{thd} = 300$ and normalized distance $D_{thd} = 0.02$, gray difference threshold $\theta' = 75$.

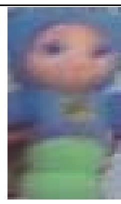
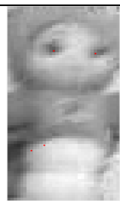
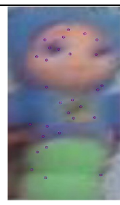
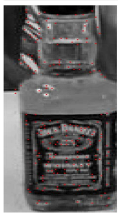
Data sequence	Target location	ORB	Improved ORB
Doll			
Liquor			

Fig.9 Schematic diagram of ORB feature detection

The target positions of two video sequences from Doll and Liquor datasets are tracked using improved ORB algorithm and ORB algorithm to compare the precision and time of feature detection. The results of the two datasets are shown in Tab.3.

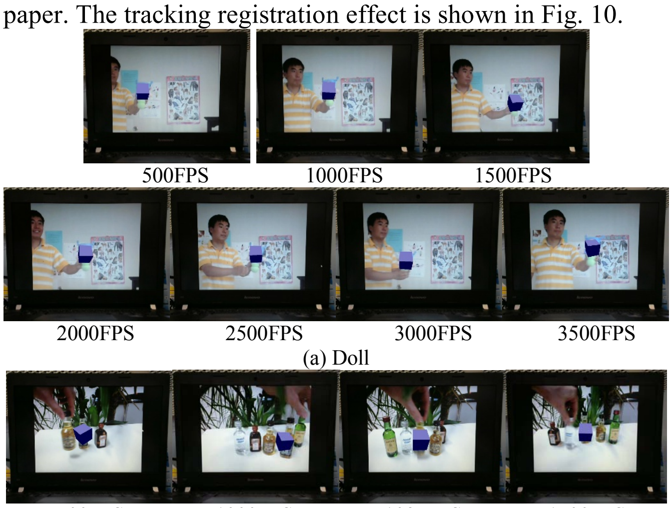
Tab.3 Comparison of feature detection precision and time

Serial	Precision/%		Time/ms	
	ORB	Improved ORB	ORB	Improved ORB
Doll	68.65	89.64	168.5	164.3
Liquor	72.30	95.62	159.4	160.2
Average	70.48	92.63	163.95	162.25

Tab.3 shows that the improved ORB algorithm improves the precision of feature detection by about 22% compared with the ORB algorithm, and the running time of the improved ORB algorithm is basically the same as that of the ORB algorithm.

C. 3D Registration Effect and Result Analysis

In this paper, the virtual information of OpenGL cube is superimposed in the real world to achieve the goal of combining virtual and real, in which all sizes of the output window of video stream are 640×480 pix. The video sequences of Doll dataset and Liquor dataset are displayed on the display, and the virtual information is superimposed on the tracking target position by the registration method in this paper. The tracking registration effect is shown in Fig. 10.



500FPS 1000FPS 1500FPS

2000FPS 2500FPS 3000FPS 3500FPS

(a) Doll

500FPS 1000FPS 1097FPS 1500FPS

(b) Liquor

Fig. 10 Long-term 3D registration based on LCT tracking and improved ORB detection

The experimental results show that the long-term 3D registration method based on LCT tracking and improved ORB detection performs well in long-term tracking and complex environments. Although Liquor dataset failed to track registration at 1000 frames, it quickly restored tracking registration after 97 frames.

The experimental results show that the long-term 3D registration method based on LCT tracking and improved ORB detection performs well in long-term tracking and complex environment. In order to further verify the precision of this 3D tracking registration method, as shown in Tab.4, in the process of tracking registration, the average and maximum errors between the tracking registration results of Doll and Liquor datasets in the x, y and z axes and the real tracking registration values are obtained.

Tab.4 3D registration error results and analysis

3D Registration method	Datasets	Average error /mm			Maximum/mm			Time /ms
		x	y	z	x	y	z	
Registration method of this article	Doll	2.0	1.5	1.8	3.2	4.1	3.8	65.63
	Liquor	1.9	1.7	1.9	4.1	2.9	4.6	
	Average	1.95	1.6	1.85	3.65	3.5	4.2	
LCT and ORB Registration method	Doll	2.6	1.7	2.3	3.0	4.8	5.1	63.89
	Liquor	2.2	1.8	2.5	3.9	3.3	8.6	
	Average	2.4	1.75	2.4	3.45	4.05	6.85	

As can be seen from Tab.4, the 3D registration method in this paper that the average error of Doll and Liquor datasets x, y and z axes is about 2 mm and the maximum error is about 3.5mm. The average error of LCT and ORB registration methods is about 2.4mm and the maximum error is about 4mm. Although the 3D registration time of this algorithm is 2ms higher than that of LCT and ORB registration methods, 2ms is permissible in practical application. Therefore, the registration method in this paper can achieve high registration accuracy. Generally, the long-term 3D registration method based on LCT tracking and improved ORB detection is more widely used in the scene, and performs better in real-time, stability and robustness.

VI. CONCLUSIONS

In this paper, we propose a long-term registration method based on LCT tracking and improved ORB detection, which makes the AR system perform well in complex environment and long-term tracking registration in 3D registration. By tracking the registered area with LCT which is more reliable in long-term tracking, setting adaptive threshold for ORB algorithm with good real-time performance, using the number and distance relationship of feature points in texture-rich regions and discrete difference features to detect the registered area, finally calculating the 3D registration matrix, completing the enhancement of the real world. The simulation results show that the 3D registration method in this paper has high reliability, high precision of feature detection and matching, and good real-time performance. The 3D registration method in AR has better practicability, stability and robustness. Because LCT algorithm is not real-time in long-term target tracking and related filtering algorithms, the next step is to improve the LCT long-term tracking algorithm by using deep learning method to improve the real-time performance of registration area tracking.

ACKNOWLEDGMENT

This research was supported by the National Nature Foundation of China (Grant No. 41761082, 61162016, 61562057), the Gansu Science and Technology Project (Grant No. 18JR3RA104, 1504FKCA038), the Science and Technology Project of Gansu Education Department (Grant No. 2017D-08), Lanzhou Jiaotong University Innovation and Entrepreneurship Course Construction Project in 2019 (2019CXCYKC10), Lanzhou Jiaotong University Experimental Teaching Reform Research Project (2019026) , Lanzhou Jiaotong University Experimental Teaching Reform

Research Project (2019007) and Lanzhou Talent Innovation and Entrepreneurship Project(2015-RC-86). all support is gratefully acknowledged.

REFERENCES

[1] Qin Z , Tai Y , Xia C , et al. Towards Virtual VATS, Face, and Construct Evaluation for Peg Transfer Training of Box, VR, AR, and MR Trainer[J]. Journal of Healthcare Engineering, 2019, 2019(1):1-10.

[2] Yin Hongpeng, Chen Bo, Chai Yi, et al. Summary of Vision-based Target Detection and Tracking [J].Journal of Automation, 2016, 42(10).

[3] Chen Z, Hong Z, Tao D. An Experimental Survey on Correlation Filter-based Tracking[J]. Computer Science, 2015, 53(6025):págs. 68-83.

[4] Bolme D S , Beveridge J R , Draper B A , et al. Visual object tracking using adaptive correlation filters[C]// The Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2010, San Francisco, CA, USA, 13-18 June 2010. IEEE, 2010:2544-2550.

[5] João F. Henriques, Caseiro R , Martins P , et al. Exploiting the Circulant Structure of Tracking-by-Detection with Kernels[M]// Computer Vision – ECCV 2012. Springer Berlin Heidelberg, 2012:702-715.

[6] Henriques J F , Caseiro R , Martins P , et al. High-Speed Tracking with Kernelized Correlation Filters[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3):583-596.

[7] Danelljan M , Khan F S , Felsberg M , et al. Adaptive Color Attributes for Real-Time Visual Tracking[C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2014:1090-1097.

[8] Danelljan M , Bhat G , Khan F S , et al. ECO: Efficient Convolution Operators for Tracking[J]. CVPR, 2017：6931-6939.

[9] Santner J , Leistner C , Saffari A , et al. PROST: Parallel robust online simple tracking[C]// 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2010:723-730.

[10] Kalal Z , Mikolajczyk K , Matas J . Tracking-Learning-Detection[J]. IEEE Transactions on Software Engineering, 2011, 34(7):1409-1422.

[11] Ma C , Yang X , Zhang N C , et al. Long-term correlation tracking[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2015:5388-5396.

[12] Wang T Y, Dong W B, Wang ZH Y. Position and orientation measurement system based on monocular vision and fixed target[J]. Infrared and Laser Engineering., 2017, 46(4):146-153.

[13] Bay H, Tuytelaars T, Gool L V. SURF: speeded up robust features[J]. Computer Vision & Image Understanding, 2006, 110(3):404-417.

[14] Rublee E, Rabaud V, Konolige K, et al. ORB: An efficient alternative to SIFT or SURF[C]// International Conference on Computer Vision. IEEE Computer Society, 2011:2564-2571.

[15] E . Rosten and T. Drummond . Machine learning for highspeed corner detection[C]. In European Conference on Computer Vision,2006: 430-443.

[16] Gao S, Tan X, Huang C. Improved algorithm of image registration based on SURF[J]. Journal of PLA University of Science and Technology, 2013,14(4):372-376.

[17] Zhang Y, Zou Z. Automatic registration method for remote sensing images based on improved ORB algorithm[J]. Remote Sensing for Land and Resources, 2013.25(3):20-24.

[18] Xu C. 3D objects in AR registration method and its application research[D]. Wuhan: Huazhong University of Science and Technology, 2011.

[19] Bahraini M S , Bozorg M , Rad A B . SLAM in dynamic environments via ML-RANSAC[J]. Mechatronics, 2018, 49:105-118.

[20] Zhang Z. Flexible camera calibration by viewing a plane from unknown orientations. Proceedings of 7th ICCV, 1999, pp.666~673.

[21] Wu Y , Lim J , Yang M H . Online Object Tracking: A Benchmark[C]// 2013 IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2013:2411-2418.

[22] Babenko B, Yang M H, Belongie S. Robust object tracking with online multiple instance learning[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2011, 33(8):1619-1632.

Modification Notes:

1.The author should note the word spacing and line spacing.

The format of the paper has been modified according to the template provided by the official website.

2.The topic of the paper contains the LCT algorithm, but the innovation of the LCT algorithm is insufficient, the author should reduce the introduction of LCT algorithm.

3.The author should put the innovation of the ORB algorithm in a more central position.

4.The language of the whole paper should be refined, rather than too detailed.For example,the past work should not go into too much details.

5.In the section of Improvement of ORB algorithm,Fig.4 should be replaced by Fig.3.

After careful comparison, it is found that the chart labels in this paper are correct.

6.The author should compare the experimental results with the latest algorithms.

7. there are no descriptions about the total running time of the proposed 3D registration method. So, readers do not judge whether the proposed method meets requirements of the intended application in terms of running time.

8. there are no comparison with other methods in terms of precision of 3D registration.

In this paper, the last part of the experiment is deeply studied, and the 3D registration method is evaluated from the existing algorithms, the average of 3D registration errors, and the three-dimensional registration time.